

AD-A046 380

WISCONSIN UNIV MADISON MATHEMATICS RESEARCH CENTER
MOVING-WEIGHTED-AVERAGE SMOOTHING EXTENDED TO THE EXTREMITIES O--ETC(U)
AUG 77 T N GREVILLE
MRC-TSR-1786

F/G 12/1

DAAG29-75-C-0024

ML

UNCLASSIFIED

|OF|
AD
A046380



END

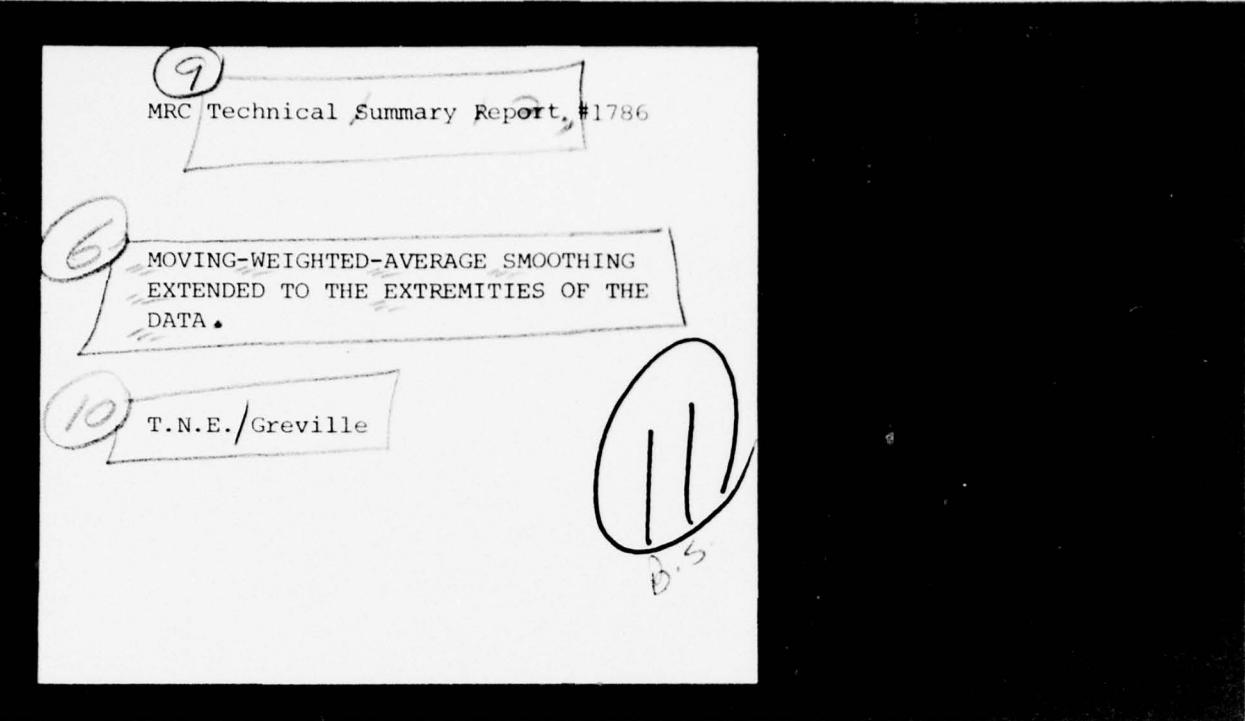
DATE

FILMED

12-77

DDC

ADAC 46380



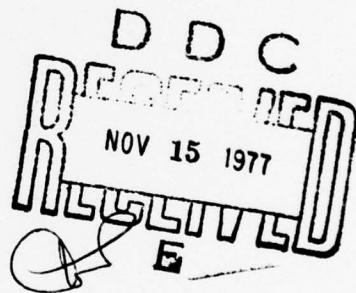
**Mathematics Research Center
University of Wisconsin-Madison
610 Walnut Street
Madison, Wisconsin 53706**

(11) August 1977

(12) 57 p.

(Received June 22, 1977)

(14) MRC-TSR-1786



(15) DAAG29-75-C-0024

**Approved for public release
Distribution unlimited**

ADAC
DDC FILE COPY:

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

221 200

4B

UNIVERSITY OF WISCONSIN - MADISON
MATHEMATICS RESEARCH CENTER

MOVING-WEIGHTED-AVERAGE SMOOTHING EXTENDED TO
THE EXTREMITIES OF THE DATA

T. N. E. Greville

Technical Summary Report # 1786
August 1977

ABSTRACT

A symmetrical moving weighted average (MWA) for smoothing observational data which may be regarded as equally spaced measurements of a function of one variable has the form

$$u_x = \sum_{j=-m}^m c_j y_{x-j} \quad (1)$$

where y_x is an observed value, u_x is the corresponding smoothed value, and the c_j are real coefficients whose sum is unity, with $c_{-j} = c_j$. This process does not yield smoothed values of the first m and the last m observations unless additional data are available. A natural method is suggested for extending the smoothing to the extremities of the data.

If (1) is exact for polynomials up to the degree $2s - 1$, it can be written in the form

$$u_x = [1 - (-1)^s \delta^{2s} q(E)] y_x ,$$

where δ is the finite difference taken centrally, E is defined by $Ef(x) = f(x + 1)$, and

$$q(E) = \sum_{j=-m+s}^{m-s} q_j E^j$$

for some coefficients q_j . If $q(z)$ has no zero on the unit circle, there is a Laurent expansion

$$[q(z)]^{-1} = \sum_{j=-\infty}^{\infty} h_j z^j$$

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

convergent in an annulus containing the unit circle.

We regard the overall smoothing process as a matrix-vector operation

$$u = Gy, \quad (2)$$

where u and y are vectors of N components and G is symmetric with rows, except for the first m and the last m , that merely reflect the application of (1). We determine the first m and the last m rows by taking

$$G = I - K^T D K,$$

where K is the matrix of $N - s$ rows and N columns that transforms a vector into the vector of s th finite differences of its components, and D is the symmetric matrix of order $N - s$ whose inverse is the Toeplitz matrix $T = (t_{ij}) = (t_{i-j})$, with $t_{i-j} = h_{i-j}$.

The same vector u can be obtained by a computational short-cut. Let $p(z)$ be the monic polynomial of degree $m - s$ whose zeros are those zeros of $q(z)$ lying within the unit circle, and let

$$a(z) = (z - 1)^s p(z) = z^m - \sum_{j=1}^m a_j z^{m-j}.$$

Then, if the range of x is from A to B , recursively calculate fictitious extended values y_x for $x = A - 1, A - 2, \dots, A - m$ by

$$y_x = \sum_{j=1}^m a_j y_{x+j}.$$

Similarly, calculate extended values for $x = B + 1, B + 2, \dots, B + m$ recursively by

$$y_x = \sum_{j=1}^m a_j y_{x-j}.$$

Finally, apply (1) to the entire sequence of observed and extended values to obtain smoothed values for $x = A, A + 1, \dots, B$.

Schoenberg (1946) defined the characteristic function of (1) as

$$\phi(t) = \sum_{j=-m}^m c_j e^{ijt},$$

and calls an MWA a smoothing formula if

$$-1 \leq \phi(t) \leq 1 ,$$

with some ambiguity as to whether the inequalities should be strict for $0 < t < 2\pi$.

It is shown here that the limit $\lim_{n \rightarrow \infty} G^n$ exists for all $N > 2m$ if and only if
 $-1 \leq \phi(t) < 1$ for $0 < t < 2\pi$.

If $0 \leq \phi(t) < 1$ for $0 < t < 2\pi$, (2) is equivalent to the minimization of

$$(u - y)^T (u - y) + (Ku)^T H Ku ,$$

where $H = (D^{-1} - KK^T)^{-1}$ is positive definite. This generalizes the Whittaker (1923) smoothing process.

SIGNIFICANCE AND EXPLANATION

The use of a moving weighted average of $2m + 1$ terms to smooth equally spaced observations of a function of one variable does not yield smoothed values of the first m and the last m observations, unless additional data beyond the range of the original observations are available. Using Toeplitz matrices, Laurent series, and analogies to the Whittaker smoothing process, we develop a natural method of extending the smoothing to the extremities of the data.

AMS(MOS) Subject Classification - 65D10, 65F30

Key Words - Smoothing, Toeplitz matrix, Laurent series, Moving weighted average

Work Unit Number 2 - Other Mathematical Methods

Sponsored by the United States Army under Contract No. DAAG29-75-C-0024.

ACCESSION for	
VIS	White Section <input checked="" type="checkbox"/>
NDG	Buff Section <input type="checkbox"/>
ANNOUNCED	
ISI LIST	
BY	
DISTRIBUTION/AVAILABILITY CODES	
SP.CIAL	
P	

1. INTRODUCTION

A time-honored method of smoothing equally spaced observations of a function of one variable to remove or reduce unwanted irregularities is the moving weighted average (MWA). An example is Spencer's 15-term average (Macaulay 1931; Henderson 1938), which can be expressed in the form

$$u_x = \frac{1}{320} (-3y_{x-7} - 6y_{x-6} - 5y_{x-5} + 3y_{x-4} + 21y_{x-3} + 46y_{x-2} + 67y_{x-1} + 74y_x \\ + 67y_{x+1} + 46y_{x+2} + 21y_{x+3} + 3y_{x+4} - 5y_{x+5} - 6y_{x+6} - 3y_{x+7}) , \quad (1.1)$$

where y_x is the observed value corresponding to the argument x , and u_x is the corresponding adjusted value. Actuarial writers commonly refer to such smoothing of data as "graduation."

More generally (Schoenberg 1946) a symmetrical MWA is of the form

$$u_x = \sum_{j=-m}^m c_j y_{x-j} , \quad (1.2)$$

where m is a given positive integer and the real coefficients c_j are such that $c_{-j} = c_j$ and

$$\sum_{j=-m}^m c_j = 1 .$$

Such averages have a long history, that includes some eminent names, but the literature concerning them is little known in the general mathematical community. Among the early writers on the subject was the Italian astronomer G. V. Schiaparelli (1866), who is chiefly remembered for his observations of the planet Mars. The majority of publications in this area have appeared in English and Scottish actuarial journals starting with John Finlaison in 1829 (see Maclean 1913). Probably the first writer to make a systematic investigation of such averages was the American mathematician E. L. De Forest (1873, 1875, 1876, 1877). His work, published in obscure places, was rescued from total oblivion largely through the efforts of Hugh H. Wolfenden (1892-1968), who also made important contributions to the subject (Wolfenden 1925). E. T. Whittaker (1923) suggested an alternative method of

smoothing, which has been widely employed, especially by actuaries, and will be referred to extensively later, because of numerous analogies to the MWA procedure. The first writer to apply sophisticated mathematical tools to the study of these averages was I. J. Schoenberg (1946, 1958, 1953), who introduced the notion of the characteristic function of an MWA, and utilized it to formulate a criterion for judging whether a given average can properly be called a "smoothing formula." This criterion will be discussed in Section 10.

2. THE PROBLEM OF SMOOTHING NEAR THE EXTREMITIES OF THE DATA

When MWA's have been used by actuaries, the argument x is usually age (of a person) in completed years. When they are used for smoothing economic time series, x denotes the position of a particular observation in a time sequence. The latter area of application appears to stem largely from the work of Frederick R. Macaulay (1931), who was the son of an actuary.

In either case, a serious disadvantage of the method is that it does not produce adjusted values for arguments too near the extremities of the data. For example, suppose Spencer's 15-term average is used to smooth monthly data extending from 1970 through 1976. The formula does not give smoothed values for the first 7 months of 1970 or the last 7 months of 1976 unless data can be obtained for the last 7 months of 1969 and the first 7 months of 1977. Clearly, acquisition of data extending farther into the past is less of a problem than acquisition of future data.

Actuaries in North America seem to have largely abandoned the use of MWA's in favor of Whittaker's method, which does not have the disadvantage described. It is likely that British actuaries may still use these averages to some extent. They appear to be currently employed by economic and demographic statisticians (Shiskin, Young, and Musgrave 1967).

Various suggestions have been made (De Forest 1877, Miller 1946, Greville 1957, 1974a) for dealing with the problem of adjustment of data near the extremities, but none of them have won general acceptance. De Forest's (1877, p. 110) suggestion is so relevant to the subject of the present paper that it is worth quoting in full:

"As the first m and the last m terms of the series cannot be reached directly by the formula [of $2m + 1$ terms], the series should be graphically extended by m terms at both ends, first plotting the observations on paper as ordinates, and then extending

the curve along what seems to be its probable course, and measuring the ordinates of the extended portions. It is not necessary that this extension should coincide with what would be the true course of the curve in those parts. The important part is that the m terms thus added, taken together with the $m + 1$ adjacent given terms, should follow a curve whose form is approximately algebraic and of a degree not higher than the third."

Elsewhere (Greville 1974a) I have proposed extrapolating the observed data by fitting a least-squares cubic to the first $m + 1$ values and a similar cubic to the last $m + 1$ observations. This is very much in the spirit of De Forest's suggestion; it is not a long step from graphic to algebraic extrapolation.

Another approach (Greville 1957) regards the adjustment process as a matrix-vector operation. We write

$$u = Gy ,$$

where y is the vector of observed values, u is the corresponding vector of adjusted values, and G is a square matrix. If a specified symmetrical MWA of $2m + 1$ terms is to be used wherever possible, then the nonzero elements of G , except for the first m and the last m rows, are merely the weights in the moving average, these weights moving to the right as one proceeds down the rows of the matrix. In the first m and the last m rows special unsymmetrical weights, determined in some appropriate manner, must be inserted. The matrix approach and the extrapolation are not wholly unrelated, since the final results of the extrapolation approach can be expressed in matrix form.

It is the purpose of the present paper to show that when a given MWA is being employed, there is a natural, preferred method of extending the adjustment to the extremities of the data, strongly suggested by the mathematical properties of the weighted average. This natural method of extension seems to have eluded previous writers on the subject, as indeed it eluded me during the many years I have thought about the matter. The preferred method of extension has the interesting property that it can be arrived at either through the matrix approach or the extrapolation approach. In the latter case, one must employ a special extrapolation formula uniquely determined by the given MWA. Though the two approaches appear to be quite different, they will be shown in Section 9 to be mathematically equivalent, and they will give identical results except for rounding error.

In my own thinking I arrived at the procedure first through the matrix approach, guided largely by extensive analogies to the Whittaker process (which is most conveniently expressed in matrix terms). It was only later that I became aware that identical results could be obtained by means of an extrapolation algorithm. Though the matrix approach provides far greater insight into the rationale behind the procedure, the extrapolation approach is simpler computationally. Therefore, we shall first describe and illustrate the extrapolation algorithm, and shall then motivate and justify the procedure by means of the matrix approach.

The extrapolation approach is merely a computational short cut, and nearly always the extended values obtained by its use are highly unrealistic if regarded as extrapolated values of the function under observation. This fact is irrelevant, but has seriously "turned off" some users. Hereafter I shall therefore avoid the use of the words "extrapolate" and "extrapolation," and shall speak of "extension," "extended values," and "intermediate values."

It is emphasized that the procedure to be described (or any other procedure for completing the graduation) is recommended for use only when additional data extending beyond the range of the original data are not available.

3. THE EXTENSION ALGORITHM

A weighted average of the form (1.2) will be called exact for the degree r if it has the property that, in case all the observed values y_{x-j} in (1.2) should happen to be the corresponding ordinates of some polynomial $P(x - j)$ of degree r or less, then

$$u_x = y_x = P(x) . \quad (3.1)$$

In other words, an average that is exact for the degree r reproduces without change polynomials of degree r or less. If the weights are symmetrical, r must be odd, and we may write $r = 2s - 1$. This implies that $r < 2m + 1$, and therefore $s \leq m$.

For a simple (unweighted) average, $r = 1$. For the overwhelming majority of MWA's used in practice, $r = 3$. The preference for cubics has a long history. De Forest (1873, p. 281) suggests that "a curve of the third degree, which admits a point of inflexion ... is ... better adapted than the common parabola to represent the form of a series whose second difference changes its sign."

We shall use the notation of the calculus of finite differences, wherein E is the "displacement operator" defined by

$$Ef(x) = f(x + 1),$$

and δ is the "central difference" operator defined by

$$\delta f(x) = f(x + \frac{1}{2}) - f(x - \frac{1}{2}),$$

so that

$$\delta^2 f(x) = f(x + 1) - 2f(x) + f(x - 1).$$

If the weighted average (1.2) is exact for the degree $2s - 1$, it can be written in the form

$$u_x = [1 - (-1)^s \delta^{2s} q(E)] y_x, \quad (3.2)$$

where $q(E)$ is of the form

$$q(E) = \sum_{j=-m+s}^{m-s} q_j E^j \quad (3.3)$$

with $q_{-j} = q_j$. In a typical smoothing formula $q(E)$ has only positive coefficients, but this is not necessarily the case. If $q(z)$ is multiplied by z^{m-s} to eliminate negative exponents, the resulting polynomial is of degree $2m - 2s$. Because of the symmetry of the coefficients, it is a reciprocal polynomial. In other words, if r is a zero of the polynomial, it follows that r^{-1} is a zero. We shall make the assumption that this polynomial has no zero on the unit circle. If it does have such zeros, the extension of the smoothing process to the extremities of the data is undefined.

Let $p(z)$ denote the polynomial of degree $m - s$ with leading coefficient unity whose zeros are the $m - s$ zeros of $z^{m-s} q(z)$ located within the unit circle. In general, some or all of these zeros are complex, but they must occur in conjugate pairs, so that $p(z)$ has real coefficients. Now we define a polynomial $a(z)$ of degree m and its coefficients a_j by

$$a(z) = (z - 1)^s p(z) = z^m - \sum_{j=1}^m a_j z^{m-j}. \quad (3.4)$$

Suppose the given data consist of $N = B - A + 1$ given values extending from $x = A$ to $x = B$. We assume that $N \geq 2m + 1$, so that at least one smoothed value is obtained

by direct application of the given MWA. Then we obtain m intermediate values to the left of $x = A$ by successive application of the recurrence

$$y_x = \sum_{j=1}^m a_j y_{x+j} . \quad (3.5)$$

Similarly, m intermediate values to the right of $x = B$ will be obtained by the analogous recurrence

$$y_x = \sum_{j=1}^m a_j y_{x-j} .$$

Finally, application of the symmetrical MWA of $2m + 1$ terms to the $N + 2m$ observed and intermediate values gives adjusted values u_x for $x = A, A + 1, \dots, B$.

For example, Spencer's 15-term formula (1.1) can be expressed in the form (3.2) with $s = 2$, where

$$\begin{aligned} q(E) = \frac{1}{320} & (3E^{-5} + 18E^{-4} + 59E^{-3} + 137E^{-2} + 242E^{-1} + 318 + 242E + 137E^2 \\ & + 59E^3 + 18E^4 + 3E^5) . \end{aligned}$$

Using a computer program to find the zeros of $z^5 q(z)$, constructing the polynomial $p(z)$, and finally applying the formula (3.4), we obtain for Spencer's 15-term formula

$$\begin{aligned} a(z) = z^7 - .961572z^6 - .372752z^5 - .015904z^4 + .123488z^3 + .125229z^2 \\ + .075887z + .025624 . \end{aligned}$$

The coefficients are rounded to the nearest sixth decimal place, except that the final digits of the coefficients of z^3 and z^2 have been adjusted by one unit to make the sum of the coefficients exactly zero.

Note that in the trivial case $s = m$, $q(z)$ is a constant and $p(z)$ is unity. Thus the algorithm reduces to extrapolation of the observed data by s th differences (i.e., by fitting a polynomial of degree $s - 1$ to the first s observations).

As a numerical illustration, Spencer's 15-term average has been applied to some meteorological data. Table 1 and Figure A show the observed and graduated values of monthly precipitation in Madison, Wisconsin in the years 1967-71. No adjustment has been made for the unequal length of the months.

1. Monthly Precipitation (Inches), Madison, Wisconsin, 1967-71.

<u>Year and Month</u>	<u>Observed Value</u>	<u>Graduated Value</u>	<u>Year and Month</u>	<u>Observed Value</u>	<u>Graduated Value</u>	
1967	January	1.63	1.11	1969	July	4.28
	February	1.17	1.63		August	0.96
	March	1.49	2.24		September	1.35
	April	2.57	2.88		October	2.65
	May	3.53	3.42		November	0.70
	June	6.46	3.74		December	1.66
	July	2.51	3.85		January	0.44
	August	2.71	3.75		February	0.16
	September	2.68	3.42		March	1.17
	October	5.52	2.92		April	2.53
	November	1.83	2.31		May	6.09
	December	1.89	1.69		June	2.26
1968	January	0.56	1.31		July	2.42
	February	0.49	1.36		August	0.97
	March	0.59	1.87		September	8.82
	April	4.18	2.69		October	2.65
	May	2.02	3.49		November	1.06
	June	7.82	3.91		December	2.12
	July	2.54	3.92		January	1.48
	August	2.58	3.54		February	2.59
	September	4.45	2.97		March	1.52
	October	0.85	2.45		April	2.42
	November	1.74	1.99		May	0.98
	December	2.89	1.64		June	2.27
1969	January	2.26	1.56		July	1.65
	February	0.18	1.81		August	3.96
	March	1.47	2.35		September	1.87
	April	2.72	3.13		October	1.30
	May	3.45	3.81		November	3.48
	June	7.96	4.05		December	3.64
1970						
1971						

SOURCE: Observed values from U. S. Department of Commerce, National

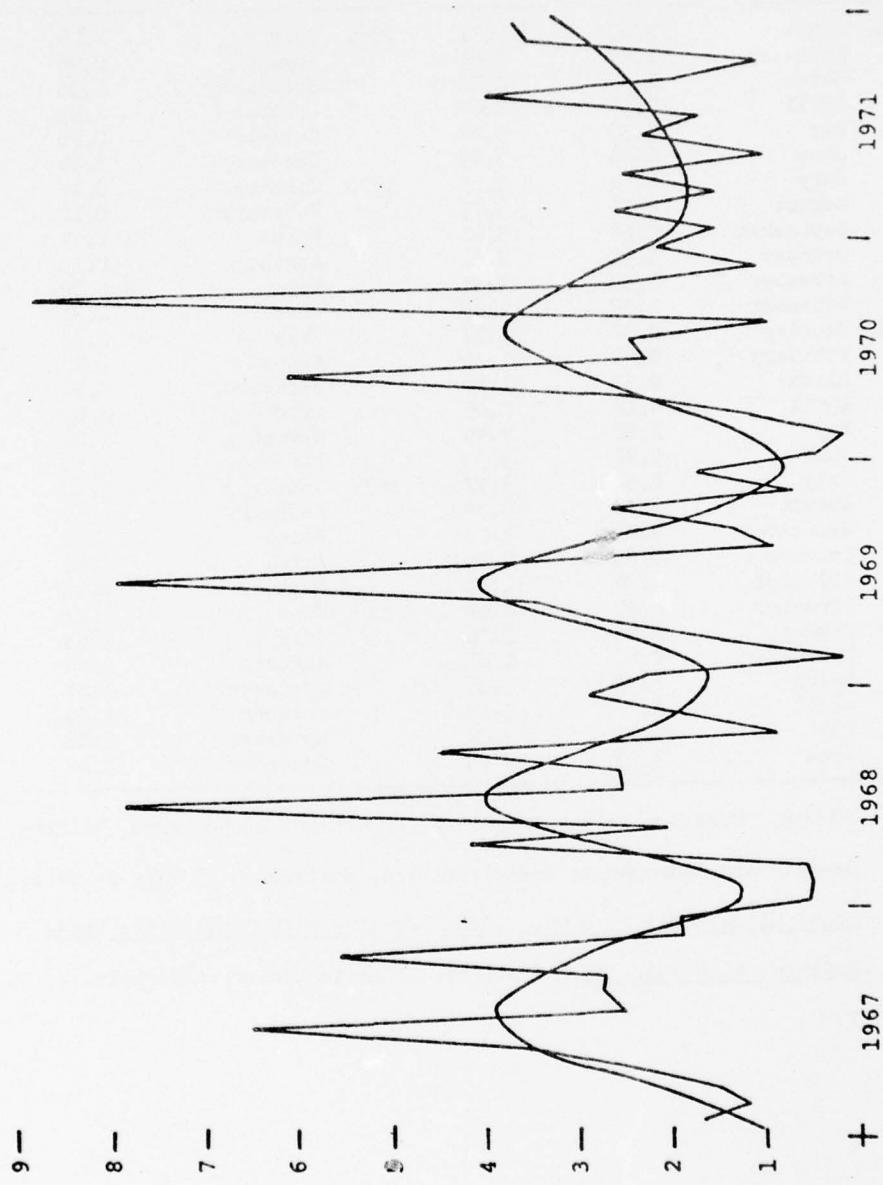
Oceanic and Atmospheric Administration, Environmental Data Service,

Local Climatological Data, Annual Summary with Comparative Data.

Madison, Wisconsin, 1972, National Climatic Center, Asheville, N. C.,

1973.

A. Observed and Graduated Values of Monthly
Precipitation, Madison, Wisconsin, 1967-1971



For the convenience of the user, the weighted-average coefficients and the intermediate-value coefficients for those averages that appear to be in common use or are found in publications accessible to me are given in the next section in Tables 2 and 3. The reader who is more interested in the justification of the procedure and the rationale behind it may skip at once to Section 5.

4. TABLES OF MOVING-AVERAGE AND EXTENSION COEFFICIENTS

Tables 2 and 3 show the coefficients in the MWA and the corresponding extension coefficients (that is, c_j and a_j) for 21 weighted averages that have appeared in the literature. Table 2 is devoted to the class of averages known to actuaries as minimum- R_3 formulas and to economic statisticians as "Henderson's ideal" formulas. They are discussed more fully in Section 7. The values in Table 2 are shown to six decimal places. In both instances, a few final digits have been adjusted by one unit to make the sum exactly unity. The moving-average coefficients are given to the nearest sixth decimal place except for the slight adjustments mentioned; rounding error in the computation of the extension coefficients may have introduced further small errors in some instances.

Table 3 is concerned with 11 moving averages derived by various writers on an ad hoc basis and known by the names of their originators. The source notes for this table do not attempt to cite the earliest publication of the formula in question, but merely indicate a convenient reference where it can be found. All these averages are exact for cubics except Hardy's, which is exact only for linear functions. The coefficients in the averages of Table 3 are rational fractions with relatively small denominators, and the user will probably find it convenient to use as weights the integers in the numerators of the coefficients, dividing by the common denominator as the final step. The column headings, therefore, are c_j multiplied by the common denominator.

In both Tables 2 and 3 advantage has been taken of the symmetry of the coefficients c_j to reduce the length of the columns by approximately one-half. The manner of using the tables may be illustrated by taking Spencer's 15-term average as an example. Equation (1.1) shows the calculation of the moving averages. The intermediate values y_x for $x = A - 1$ to $A - 7$ are calculated successively by the formula

$$y_x = .961572y_{x+1} + .372752y_{x+2} + .015904y_{x+3} - .123488y_{x+4} - .125229y_{x+5} \\ - .075887y_{x+6} - .025624y_{x+7} .$$

The intermediate values for $x = B + 1$ to $B + 7$ are calculated by the identical formula except that the "+" signs in the subscripts are changed to "-" signs.

The extension procedure drastically reduces the number of values that need to be tabulated for a given weighted average, and makes it possible, for example, to give complete information about 21 such averages in the reasonably compact Tables 2 and 3. However, the user who intends to apply a single weighted average to many data sets may prefer to tabulate the atypical elements of the smoothing matrix G for that weighted average, and so avoid the extra step of calculating the intermediate values. For the benefit of such users, a method of calculating the atypical rows of G will now be described. Justification of the procedure will be given in Section 9 (see equation (9.10)). We observe that the nonzero elements in each row of G except the first m and the last m rows are merely the coefficients c_j of the MWA centered about the diagonal element. The elements in the first m rows of G , except for the first m columns, follow from the symmetry of G , and if $G = (g_{ij})$, we have

$$g_{ij} = c_{j-i} .$$

This leaves only the square submatrix of order m in the upper left corner to be calculated. Let c denote the constant $-q_{m-s}/p_{m-s}$, where $-p_{m-s}$ is the term free of z in the polynomial $p(z)$, and let $A_1 = (a_{ij})$ denote the square matrix of order m given by

$$a_{ij} = \begin{cases} 0 & \text{for } i > j \\ 1 & \text{for } i = j \\ -a_{j-i} & \text{for } i < j . \end{cases}$$

Then the required submatrix in the upper left corner of G is given by

$$I - c A_1^T A_1 .$$

The similar submatrix in the lower right corner of G contains the same elements, but with the order of both rows and columns reversed.

2. Moving-Average Coefficients (c_j) and Extension
 Coefficients (a_j) of Minimum- R_3 ("Henderson's Ideal")
 Averages of 5 to 23 Terms Exact for Cubics

j	Number of Terms									
	5		7		9		11		13	
	c_j^a	a_j	c_j^a	a_j	c_j^a	a_j	c_j^a	a_j	c_j^a	a_j
0	.559440		.412588		.331140		.277944		.240058	
1	.293706	2	.293706	1.618034	.266557	1.352613	.238693	1.160811	.214337	1.016301
2	-.073426	-1	.058741	-.236068	.118470	,114696	.141268	.281079	.147356	.360880
3			-.058741	-.381966	-.009873	-.287231	.035723	-.140968	.065492	-.021625
4					-.040724	-.180078	-.026792	-.204545	0	-.160909
5							-.027864	-.096377	-.027864	-.138330
6									-.019350	-.056317

^aCalculated by formula (7.5).

2. Moving-Average Coefficients (c_j) and Extension
 Coefficients (a_j) of Minimum- R_3 ("Henderson's Ideal")
 Averages of 5 to 23 Terms Exact for Cubics (continued)

j	Number of Terms									
	15		17		19		21		23	
	c_j^a	a_j	c_j^a	a_j	c_j^a	a_j	c_j^a	a_j	c_j^a	a_j
0	.211542		.189232		.171266		.156470		.144060	
1	.193742	.903661	.176390	.813444	.161691	.739580	.149136	.678000	.138318	.625880
2	.145904	.397295	.141112	.410885	.134965	.412090	.128423	.406495	.121949	.397207
3	.082918	.064751	.092293	.124932	.096658	.166162	.097956	.193174	.097395	.212501
4	.024028	-.100710	.042093	-.043456	.054685	.005097	.063038	.046016	.068303	.075236
5	-.014134	-.135445	.002467	-.110644	.017474	-.078255	.029628	-.046290	.038933	-.015313
6	-.024499	-.094424	-.018640	-.106213	-.008155	-.099972	.003119	-.084020	.013430	-.063927
7	-.013730	-.035128	-.020370	-.065896	-.018972	-.081843	-.012896	-.084711	-.004948	-.078737
8										
9										
10										
11										

*Calculated by formula (7.5).

3. Moving-Average Coefficients (c_j) and Extension
Coefficients (a_j) of Selected Moving Averages

j	Spencer											
	Macaulay ^a	15-Term ^b	Woolhouse ^c	125c _j	a _j	120c _j	a _j	125c _j	a _j	625c _j	a _j	
0	864c _j	a _j	320c _j	a _j	125c _j	a _j	120c _j	a _j	125c _j	a _j	125	
1	182		74		25		24		25		125	
2	171	.919760	67	.961572	24	.885108	22	.739988	24	.859550	114	.820240
3	127	.393023	46	.372752	21	.421982	17	.386211	18	.399283	87	.402924
4	72	.055273	21	.015904	7	.028721	10	.124325	10	.087040	53	.114622
5	17	-.113111	3	-.123488	3	-.076050	4	-.023648	3	-.072738	21	-.047133
6	-17	-.140462	-5	-.125229	0	-.107285	0	-.080087	0	-.104527	0	-.102491
7	-19	-.084512	-6	-.075887	-2	-.092723	-2	-.079459	-2	-.093953	-8	-.091791
8	-10	-.029971	-3	-.025624	-3	-.059753	-2	-.049327	-2	-.055312	-9	-.060239
9							-1	-.018003	-1	-.019343	-6	-.028636
											-2	-.007496

^aMacaulay 1931, p. 55, footnote 2.

^bMacaulay 1931, p. 55; Henderson 1938, p. 53.

^cHenderson 1938, p. 53.

^dHenderson 1938, p. 53; Benjamin and Haycocks 1970, p. 238.

^eHenderson 1938, p. 53.

^fHenderson 1938, p. 53.

3. Moving-Average Coefficients (c_j) and Extension
Coefficients (a_j) of Selected Moving Averages (continued)

	Andrews ^g		Spencer ^h 21-Term		Hardy Wave-Cutting ⁱ		Vaughan Formula A ^j		Kenchington ^k	
	$10080c_j$	a_j	$350c_j$	a_j	$65c_j$	a_j	$1440c_j$	a_j	$385c_j$	a_j
0	1688		60		5		182		45	
1	1579	.700747	57	.729724	5	.480996	179	.593256	44	.527740
2	1325	.406808	47	.408707	6	.368708	170	.396409	41	.370688
3	950	.179749	33	.167281	7	.267940	149	.230238	36	.236445
4	551	.027155	18	.009255	7	.166506	115	.096761	30	.128638
5	225	-.054586	6	-.069703	6	.072964	72	-.000857	22	.043118
6	-4	-.083701	-2	-.091513	4	-.008222	29	-.060076	13	-.018390
7	-124	-.078256	-5	-.076165	1	-.075454	-5	-.083321	5	-.053902
8	-135	-.054368	-5	-.049051	-1	-.097387	-26	-.079596	-1	-.067080
9	-110	-.031120	-3	-.022502	-2	-.089039	-29	-.056662	-5	-.064844
10	-61	-.012428	-1	-.006033	-2	-.062016	-19	-.028557	-6	-.050323
11					-1	-.024996	-6	-.007595	-5	-.032035
12									-3	-.015626
13									-1	-.004429

^g Andrews and Nesbitt 1965, p. 18.

^h Macaulay 1931, p. 51; Henderson 1938, p. 53.

ⁱ Benjamin and Haycocks 1970, p. 239.

^j Vaughan 1933, p. 437.

^k Henderson 1938, p. 53.

5. THE WHITTAKER GRADUATION PROCESS

It is not the purpose of this paper to consider the Whittaker (1923; see also Henderson 1924) graduation process in detail. However, since the natural method of extension of MWA graduation to the extremities of the data was arrived at primarily on the basis of analogies to the Whittaker method, the latter must be described sufficiently to make these analogies clear. The objective of the Whittaker process is to choose graduated values u_j ($j = A, A + 1, \dots, B$) in such a way as to minimize the quantity

$$\sum_{j=A}^B w_j (u_j - y_j)^2 + g \sum_{j=A}^{B-s} (\Delta^s u_j)^2 , \quad (5.1)$$

where the weights w_j , the positive constant g , and the positive integer s are chosen a priori by the user. The solution is most conveniently expressed in matrix notation as follows (Greville 1957, 1974a). Let W denote the diagonal matrix of order N whose successive diagonal elements are the w_j , let u and y be defined as in Section 2, and let K denote the rectangular matrix of $N - s$ rows and N columns that transforms a vector v into the vector of s th finite differences of its components. Clearly the non-zero elements of K are binomial coefficients of order s with alternating signs (Greville 1974a). Then, the expression (5.1) can be written in the form

$$(u - y)^T W(u - y) + g(Ku)^T Ku , \quad (5.2)$$

where the superscript T denotes the transpose. It is easily seen (Greville 1974a) that (5.2) is smallest when u satisfies

$$(W + gK^T K)u = Wy . \quad (5.3)$$

It is not difficult to show (Greville 1957, 1974a) that the matrix in the left member of (5.3) is nonsingular (in fact, positive definite) and therefore

$$u = (W + gK^T K)^{-1} Wy .$$

The remaining discussion will be limited to the so-called "Type A" case, in which all the weights w_j are taken equal to unity, as this case has the greatest similarity to MWA graduation. Here $W = I$ (the identity), and it is easily verified (Noble 1969, p. 147) that

$$(I + gK^T K)^{-1} = I - K^T (g^{-1} I + KK^T)^{-1} K . \quad (5.4)$$

If the entire process of graduation, by whatever method or criterion, including data near the ends, is conceived in terms of matrix-vector multiplication (Greville 1957), so that

$$u = Gy \quad (5.5)$$

for some matrix G , (5.4) suggests that it may be reasonable to consider matrices G of the form

$$G = I - K^T DK \quad (5.6)$$

for some square matrix D and some order of differences s .

6. MATRIX DEVELOPMENT OF THE NATURAL METHOD OF COMPLETING THE GRADUATION

We suppose that N equally spaced observed values y_j ($j = A, A + 1, \dots, B$) are to be graduated primarily by means of a given symmetrical MWA of $2m + 1$ terms of the form (1.2), that is exact for the degree $2s - 1$. We assume that $N > 2m$. In other words, graduated values u_j for $j = A + m, A + m + 1, \dots, B - m$ will be calculated from the given weighted average. This requirement fixes the elements of the matrix G of (5.5) and (5.6) with the exception of the first m and the last m rows. The nonzero elements of each of the remaining $N - 2m$ rows will be merely the weights in the moving average with the middle weight on the diagonal in each case.

Our determination of the elements of the first m and the last m rows of G will be based on the general requirement that these rows shall not be something extra grafted onto the main part of the matrix, but shall be an integral part of an overall matrix having a well defined structure, this structure having the greatest possible analogy to that of the corresponding matrix for the Whittaker process. We shall try to show that this general requirement leads almost inexorably to the following three assumptions about G for the MWA case:

- (i) G is symmetric and of the form (5.6);
- (ii) D is such that D^{-1} exists and is a Toeplitz matrix (to be defined presently);
- (iii) the elements of $D^{-1} = (d'_{ij})$ are given by

$$d'_{ij} = h_{i-j} \quad (6.1)$$

where $h_j = h_{-j}$ is a coefficient in the Laurent expansion

$$h(z) = [q(z)]^{-1} = \sum_{j=-\infty}^{\infty} h_j z^j, \quad (6.2)$$

convergent in an annulus containing the unit circle.

These three assumptions (together with the assumption stated in the first paragraph of this section about the rows of G other than the first m and the last m) uniquely determine G . The three assumptions require extensive discussion, explanation, and comment, on which we now embark.

Since analogy to the Whittaker process is to have the highest priority, and (5.4) shows that G for that process is clearly symmetric and of the form (5.6), these being very basic structural properties, there can be little question about assumption (i). This assumption implies that G is a diagonal band matrix of band width $2m + 1$, and its elements are now determined except for a square submatrix of order m in the upper left corner and a similar submatrix in the lower right corner. It also implies that D is symmetric and is a diagonal band matrix of band width $2m - 2s + 1$.

It may be mentioned here that there is one basic, unavoidable difference between the Whittaker process and the MWA process. This is that, while in the MWA process (with the natural extension) G is a diagonal band matrix, in the Whittaker process it is the inverse of such a matrix. In consequence of this difference, the Whittaker process is "global" (each graduated value depending on all the observed values), while MWA is "local" (each graduated value depending only on a few neighboring observed values). This distinction carries over to the related matrix D , which, in the Whittaker process, is not a diagonal band matrix but the inverse of such a matrix (of band width $2s + 1$); from (5.4),

$$D^{-1} = g^{-1} I + KK^T. \quad (6.3)$$

Assumption (i) fixes the elements of D except for those in a square submatrix of order $m - s$ in the upper left corner and a similar submatrix in the lower right corner. Reverting to the expression $q(E)$ of (3.2) and (3.3), we have $D = (d_{ij})$, with

$$d_{ij} = q_{i-j} \quad , \quad (6.4)$$

except within the two submatrices mentioned.

We define a Toeplitz matrix (see Trench 1974) as one in which all the elements on any diagonal line extending downward and to the right are equal. In other words, $T = (t_{ij})$ is a Toeplitz matrix when

$$t_{ij} = t_{i-j}$$

for all i and j .

It is easily verified that D^{-1} for the Whittaker process, given by (6.3), is a Toeplitz matrix. In fact, if $D^{-1} = (d'_{ij})$,

$$d'_{ij} = (-1)^{i-j} \binom{2s}{s+i-j} + g^{-1} \delta_{ij} \quad ,$$

where δ_{ij} is a Kronecker symbol.

Now, it is clear that the Toeplitz property is a very striking and obvious property of those matrices which possess it. Thus, in pursuit of our goal of maximum analogy between the Whittaker and MWA processes, we would wish, if at all possible, to make D^{-1} a Toeplitz matrix in the MWA case. Accordingly, let $D^{-1} = (d'_{ij})$ with $d'_{ij} = d'_{i-j}$ for all i and j . Since D is symmetric, D^{-1} is symmetric and $d'_{-j} = d'_j$. Consider the series

$$f(z) = \sum_{j=-\infty}^{\infty} d'_j z^j \quad (6.5)$$

(which may or may not converge). Because of (6.4) this series is a "reciprocal" of $q(z)$ at least in the formal sense that if $q(z)$ and $f(z)$ are formally multiplied together, the product is unity.

The latter fact does not uniquely determine the series (6.5). In order to achieve a unique determination, we invoke a further analogy between the MWA and Whittaker processes. We require that in the MWA case this series converge in some region of the complex plane. The corresponding series for the Whittaker case is finite, and therefore converges everywhere.

Now, a Laurent series like (6.5), if it converges anywhere, converges in an annulus.

Because of the symmetry of the coefficients, if it converges for $z = z_0$, it converges for $z = z_0^{-1}$. Therefore, the annulus of convergence, if it exists, contains the unit circle. Moreover, $[q(z)]^{-1}$ has a Laurent expansion (6.2) convergent in an annulus containing the unit circle if and only if $q(z)$ has no zero on the unit circle.

Thus, assumption (iii) is the only possible assumption consistent with assumptions (i) and (ii) that satisfies the requirement that (6.5) converge in some part of the plane, and assumption (iii) implies that $q(z)$ has no zeros on the unit circle. The prohibition against such zeros of $q(z)$ was previously alluded to in Section 3, and further reasons for insisting on it will be given in Section 10.

In reality, the part of assumption (ii) that asserts the nonsingularity of D is redundant, because it is shown in Section 9 that if a Toeplitz matrix $T (= D^{-1})$ is constructed in accordance with assumption (iii), then the square submatrices of order $m - s$ in the upper left and lower right corners of D can be chosen so that $DT = I$.

In the typical case D is a matrix of nonnegative elements (this is true in the Whittaker case), but this is not a requirement. (It is not true of Hardy's formula.)

The matrix-vector formulation does not lead at once to a convenient method for calculating the graduated values near the ends of the data. It will be shown in Section 9 to be equivalent to the extension algorithm described in Section 3, and also to the method of calculating the atypical elements of G described in Section 4.

7. SPECIAL CLASSES OF MOVING AVERAGES

Of particular interest are those moving averages known to actuaries as minimum- R_3 formulas and to economic statisticians as "Henderson's ideal" formulas. For a given number of terms $2m + 1$, this is the average (1.2), exact for the third degree, for which the quantity

$$\sum_{j=-m-3}^m (\Delta^3 c_j)^2 \quad (7.1)$$

is smallest (with the understanding that $c_j = 0$ for $|j| > m$). The "smoothing coefficient" R_3 is defined as the quantity obtained by dividing (7.1) by 20 and taking the square root. The divisor 20 is chosen because this is the value of (7.1) for the trivial case of (1.2) in which $c_0 = 1$ and $c_j = 0$ for $j \neq 0$.

The rationale for minimizing (7.1) may be explained as follows (Greville 1974a). If, for some x , u_x , u_{x+1} , u_{x+2} , and u_{x+3} are given by (1.2), which is the case for $x = A + m$ to $B - m - 3$, inclusive, then

$$\Delta^3 u_x = - \sum_{j=-m-3}^m (\Delta^3 c_j) y_{x+j+3} . \quad (7.2)$$

It has been customary to regard the smallness (in absolute value) of the third differences of the graduated values as an indication of smoothness. Therefore (7.2) suggests that smoothness is encouraged by making the quantities $\Delta^3 c_j$ numerically small, and minimizing (7.1) is a way of doing this. The formula corresponding to (7.2) for a general order of differences is

$$\Delta^s u_x = (-1)^s \sum_{j=-m-s}^m (\Delta^s c_j) y_{x+j+s} , \quad (7.3)$$

and the general formula for R_s is

$$R_s^2 = \sum_{j=-m-s}^m (\Delta^s c_j)^2 / {s \choose s} . \quad (7.4)$$

There is some question whether Henderson's contribution warrants attaching his name to the "ideal" weighted averages. De Forest (1873) treated extensively the formulas that minimize R_4 . The concept of choosing the coefficients c_j in order to minimize R_3 seems to have been first mentioned by G. F. Hardy (1909). These averages were fully discussed by Sheppard (1913) slightly earlier than by Henderson (1916). However, Henderson does seem to have been the first to give an explicit formula for the coefficient c_j in the weighted average minimizing R_3 (Henderson 1916, p. 43; Macaulay 1931, p. 54; Henderson 1938, p. 60; Miller 1946, p. 71; Greville 1974a, p. 18). If we write $k = m + 2$, so that the weighted average has $2k - 3$ terms, the formula is

$$c_j = \frac{315[(k-1)^2 - j^2](k^2 - j^2)[(k+1)^2 - j^2](3k^2 - 16 - 11j^2)}{8k(k^2 - 1)(4k^2 - 1)(4k^2 - 9)(4k^2 - 25)} . \quad (7.5)$$

Weighted averages that minimize R_s have been discussed from other points of view by Wolfenden (1925), Schoenberg (1946), and Greville (1966, 1974b).

Also deserving of special mention are the averages (exact for cubics) that minimize

R_0 , sometimes called "formulas of maximum weight" or "Sheppard's ideal" formulas. These are sometimes applied to physical measurements when the errors of observation can be regarded as random "white noise" (see discussion of "reduction of error" in Section 8). The weights are given by

$$c_j = \frac{3(3m^2 + 3m - 1) - 15j^2}{(2m - 1)(2m + 1)(2m + 3)} .$$

Weighting coefficients (c_j) and extension coefficients (a_j) for minimum- R_3 (Henderson's ideal) averages of 5, 7, ..., 23 terms are given in Table 2.

8. COMPARISON WITH OTHER METHODS. PRACTICAL CONSIDERATIONS

If a symmetrical MWA exact for the degree $2s - 1$ is being used to smooth the main part of the data, it can easily be deduced, either from the extension algorithm described in Section 3 or from the matrix formulation of (5.5) and (5.6) that the unsymmetrical weightings proposed for smoothing the first m and the last m observations are exact only for the degree $s - 1$. For example, all the averages represented in Tables 2 and 3 with the exception of Hardy's are exact for cubics, and therefore their extensions to values near the ends are exact only for linear functions. Hardy's weighted average is exact for linear functions and its extension only for constants.

The Whittaker process has a similar property. At a sufficient distance from the ends of the data, polynomials of degree $2s - 1$ are "almost" reproduced by that process. In support of this rather loose statement the following heuristic argument is advanced. For the Whittaker process

$$G = (I + gK^T K)^{-1} = I - gGK^T K .$$

Thus, if y is the vector of observed values, the vector of corrections to these values is

$$- gGK^T Ky .$$

Now, the nonzero elements of $K^T K$, with the exception of the first s and the last s rows, are binomial coefficients of order $2s$ with alternating signs. Therefore the components of $K^T Ky$, except for the first s and the last s , are $(2s)$ th differences of those of y (or their negatives if s is odd). Thus, if y is a vector of ordinates of

a polynomial of degree $2s - 1$, $K^T Ky$ is a vector of zeros except for the first s and the last s components. The components of $GK^T Ky$ are graduated values of those of $K^T Ky$, and therefore should be very small at some distance from the extremities of the data. Finally, multiplication by g , even though g is typically large, should give small corrections at a sufficient distance from the ends of the data.

Some users may consider the reduction in degree of exactness near the ends of the data a disadvantage of the natural method of extension. Before I became aware of the natural method, I had proposed (Greville 1974a) a different method of extension (already mentioned in Section 2) that does not have this particular disadvantage (though it has other shortcomings). This involves extrapolation by a polynomial of degree $2s - 1$ fitted by least squares to the first $m + 1$ observations. A similar polynomial is fitted to the last $m + 1$ observations for extrapolation at the other end of the data. There may be a gain in simplicity in using a single method of extrapolation for all symmetrical weighted averages, the particular extrapolated values depending only on the number of terms in the main formula. However, there is a loss in that the extension method is no longer tailored to the particular symmetrical average used.

Like the natural method of extension, the method using extrapolation by least squares can be collapsed into a single matrix G . When this is done, the diagonal band character of the smoothing matrix is maintained, but the symmetry is lost. Though the matrix approach is less convenient for computational purposes, the differences between the two methods are best elucidated by comparing the first m rows of the respective matrices G . This is done in Tables 4 and 5 for the case of the 9-term "ideal" formula. Here $m = 4$, but for convenience the fifth row is also shown. Its elements would be repeated in the subsequent rows, moving successively to the right, until we come to the last four rows. While an average of as few as 9 terms would seldom be used in practice, this is a convenient illustration.

As previously indicated, the first m rows and the last m rows of G may be regarded as exhibiting unsymmetrical weighted averages which are to be used near the ends of the data to supplement the symmetrical average used elsewhere. The coefficients that appear in the last m rows are the same as those in the first m rows, but the order is

reversed, both horizontally and vertically. It should be noted that the coefficients in the supplemental averages depend only on those of the underlying symmetrical average. They do not depend on N , the number of observations in the data set (which is the order of G).

The coefficients in the supplemental weighted averages based on least-square extrapolation, exhibited in Table 5, show two undesirable features. These are negative coefficients of substantial numerical magnitude, and successive waves of positive and negative coefficients as one proceeds from left to right along the rows. The number of such waves would increase as the number of terms in the underlying formula increases.

In striking contrast is the character of the coefficients of the natural extension. Like the coefficients in the underlying symmetrical formula, each row exhibits a peak in the vicinity of the main diagonal of the matrix, tapering off to a single group of negative coefficients of reduced size near the edge of the diagonal band.

In the least-squares method only a very small correction is made to the initial observed value. The corresponding correction in the natural method is more substantial.

The "second-difference correction" is the coefficient of the second-difference term when the formula is expressed in terms of increasing orders of differences in the form

$$u_x = y_x + c \Delta^2 y_{x-h} + \dots .$$

The coefficient c does not depend on the subscript $x - h$, in which there is some freedom of choice. For the formulas based on least-squares extrapolation, which are exact for cubics, the fourth-difference correction is similarly defined.

Some writers (Miller 1946, Wolfenden 1942, Greville 1974a) have regarded the observed values y_x as the sum of "true" values U_x and superimposed random errors e_x . If it is assumed that the errors e_x for different x are uncorrelated, and have zero mean and constant variance σ^2 for all x , then the variance of the error in the smoothed value u_x is $R_0^2 \sigma^2$, where R_0^2 is obtained by taking $s = 0$ in (7.4). Thus, R_0 may be interpreted as the ratio of reduction in the standard deviation of error that results from application of the weighted average.

While the assumptions underlying the preceding analysis may be questioned, nevertheless a good case can be made that, for any weighted average, R should be less than unity. Since R_0^2 is the sum of the squares of the coefficients in the average, R_0 can

4. Matrix Elements for the Natural Extension of the 9-Term Minimum-R₃ Smoothing Formula,
with Second-Difference Correction and R₀ Value for Each Supplemental Formula

		j								
		j							Second-Difference Correction	
i	1	2	3	4	5	6	7	8	9	R ₀
1	.773854	.305888	.025938	-.064956	-.040724	0	0	0	0	-.4133 .8360
2	.305888	.360106	.270804	.113799	-.009873	-.040724	0	0	0	.1457 .5579
3	.025938	.270804	.357131	.278254	.118470	-.009873	-.040724	0	0	.1931 .5429
4	-.064956	.113799	.278254	.338473	.266557	.118470	-.009873	-.040724	0	.0744 .5441
5	-.040724	-.009873	.118470	.266557	.331140	.266557	.118470	-.009873	-.040724	0 .5322

5. Matrix Elements for the Least-Squares Extension of the 9-Term Minimum-R₃ Smoothing
Formula with Fourth Difference Correction and R₀ Value for Each Supplemental Formula

		j								
		j							Fourth-Difference Correction	
i	1	2	3	4	5	6	7	8	9	R ₀
1	.985350	.058600	-.087900	.058600	-.014650	0	0	0	0	-.01465 .9928
2	.025386	.857731	.315214	-.345889	.188282	-.040724	0	0	0	-.01534 .9962
3	-.206335	.652571	.412341	.048375	.240395	-.009873	-.040724	0	0	-.41580 .8369
4	-.140189	.232497	.299136	.241547	.299136	.118470	-.009873	-.040724	0	-.68194 .5717
5	-.040724	-.009873	.118470	.266557	.331140	.266557	.118470	-.009873	-.040724	-.75525 .5322

never be less than the maximum of the absolute values of the coefficients. Thus, an average cannot be considered satisfactory if the absolute value of any coefficient is equal to or greater than unity.

As indicated in Section 7, it has long been customary to regard a graduation as smooth if the third differences of the graduated values are small in absolute value. If $G = (g_{ij})$, we have

$$u_{A+i-1} = \sum_{j=1}^N g_{ij} y_{A+j-1}$$

and therefore

$$\Delta^s u_{A+i-1} = \sum_{j=1}^N y_{A+j-1} \Delta_i^s g_{ij} \quad (8.1)$$

where the subscript of Δ indicates that the differences are taken with respect to i (i.e., down the columns of the matrix). If one avoids the corner submatrices, the nonzero elements g_{ij} in (8.1) are merely coefficients in the underlying symmetrical average, and (8.1) reduces to (7.3). This was the rationale underlying the derivation of the minimum- R_s averages.

Of course, if G is symmetric, it makes no difference whether the differences are taken horizontally or vertically. When the symmetry of G is not assumed, care must be exercised. Many years ago (Greville 1947, 1948) I published what purported to be coefficients in supplemental averages to be used near the ends of the data in conjunction with minimum- R_3 and minimum- R_4 symmetrical averages. The symmetry of G was not assumed, and I made the error of deriving the unsymmetrical coefficients by minimizing their third differences taken horizontally. The tables in question are therefore based on an incorrect assumption. Further it may be mentioned in passing that in the 1947-8 formulation the diagonal band character was not maintained, since the supplemental averages contained the full $2m + 1$ terms.

Table 6 shows, for the natural and the least-squares extensions of the 9-term minimum- R_3 formula, those third differences of the matrix elements, taken vertically, that involve elements of the first five rows. The entries in the fifth row of Table 6 would

6. Third Differences of Matrix Elements for the Natural and Least-Squares
 Extensions of the 9-Term Minimum-R₃ Smoothing Formula

		j								
		1	2	3	4	5	6	7	8	9
Natural Extension										
1	-.000046	.075817	-.006665	-.064956	-.077748	.025917	.112299	-.040724	0	
2	.130190	.101096	.084297	-.027899	-.103248	-.077748	.025917	.112299	-.040724	
3	-.098634	.059488	.112348	.055964	-.045662	-.103248	-.077748	.025917	.112299	
4	-.057216	-.021246	.066051	.095915	.045662	-.045662	-.103248	-.077748	.025917	
5	.040724	-.112299	-.025917	.077748	.103248	.045662	-.045662	-.103248	-.077748	
Least-Squares Extension										
1	-.430376	.789377	.095655	-.709595	.157447	.025917	.112299	-.040724	0	
2	-.264548	.392618	.142871	-.257320	-.033365	-.077748	.025917	.112299	-.040724	
3	-.092060	.0333815	.119784	.091815	-.069850	-.103248	-.077748	.025917	.112299	
4	.018017	-.139944	.045169	.192841	.013083	-.045662	-.103248	-.077748	.025917	
5	.040724	-.112299	-.025917	.077748	.103248	.045662	-.045662	-.103248	-.077748	

be repeated in subsequent rows, moving successively to the right. Casual inspection of the table shows that the third differences are numerically smaller for the natural extension. All of these third differences are less than 0.14 in absolute value. Two of those for the least-squares extension exceed 0.7 in absolute value.

It is instructive to compare the natural extension with the least-squares extension for the numerical example of Section 3. Though neither extension is recommended for use when additional data are available beyond the range of the original data set, nevertheless it may be of interest, purely for purposes of illustration, to choose a numerical example in which such additional data are available, and this has been done.

Table 7 and Figures B and C show, for the first seven months of 1967 and the last seven months of 1971, the observed values of precipitation in Madison, Wisconsin, and the graduated values obtained by (i) natural extension of Spencer's 15-term average, (ii) least-squares extension of the same average, and (iii) use of additional data. It will be noted that the least-squares extension is strongly constrained toward each of the two terminal observations (January 1967 and December 1971). This may be explained by the fact that all the values y_x in (1.2) that enter into the calculation of these graduated values are included in either the $m + 1$ observations to which the least-squares cubic was fitted or the m extrapolated values obtained from the same cubic. On the other hand, the natural extension and the least-squares extension are very close together at the interface with the graduated values calculated in the standard manner. Thus, for the months of July 1967 and June 1971, all but one of the values y_x entering into the computation (1.2) are identical for the two methods.

For the months closer to the interface the graduated values obtained by introducing additional data are close to those of the natural extension. This is because the supplemental unsymmetrical averages produced by the natural extension (unlike those of the least-squares extension) give relatively small weight to the observations more remote from the one being graduated (as does the underlying symmetrical formula). For example, the values for the natural extension and those obtained by the use of additional data are indistinguishable in Figure B for April to July 1967. In the last months of 1971 the deviation is greater because the first two months of 1972 were exceptionally dry. This could not have been predicted from the data for preceding months.

7. Extension of 15-Term Spencer Graduation of Madison Precipitation
Data to First Seven and Last Seven Months by Different Methods

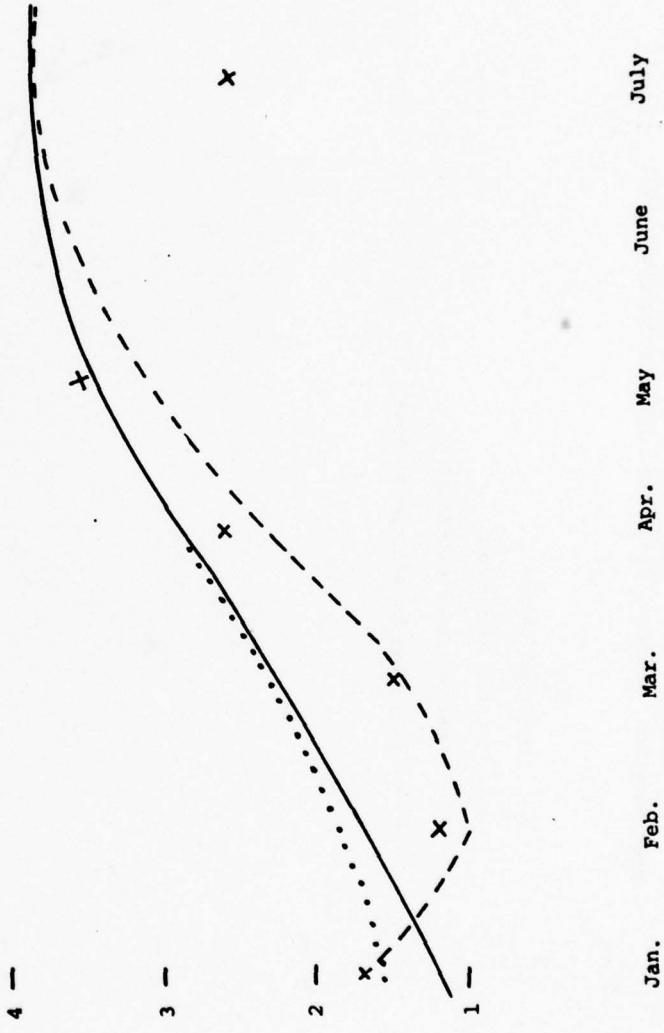
Year and Month	Extension of Graduation by			
	Observed Value	Natural Method	Least-Squares Cubic	Additional Data
1967				
January	1.63	1.11	1.62	1.56
February	1.17	1.63	0.98	1.84
March	1.49	2.24	1.37	2.29
April	2.57	2.88	2.32	2.85
May	3.53	3.42	3.07	3.36
June	6.46	3.74	3.61	3.70
July	2.51	3.85	3.82	3.84
- - - - -				
1971				
June	2.27	2.02	2.00	2.05
July	1.65	2.13	2.03	2.23
August	3.96	2.24	2.00	2.39
September	1.87	2.40	1.97	2.51
October	1.30	2.63	2.08	2.50
November	3.48	2.84	2.58	2.31
December	3.64	3.28	3.85	2.04

B. Observed and Graduated Values of Monthly
Precipitation, Madison, Wisconsin, January to
March, 1967

7 —

- Key
- 6 — X Observed values
 - Natural extension
 - Least-squares extension
 - ... Extension by additional data
 - 5 —
 - 4 —
 - 3 —
 - 2 —
 - 1 —

X



7 —

C. Observed and Graduated Values of Monthly
Precipitation, Madison, Wisconsin, July to December, 1971

Key
X Observed values
— Natural extension
- - - Least-squares extension
• • • Extension by additional data

5 —

4 — X

3 —

X

1 — X

June

Dec.

July

Oct.

Aug.

Nov.

Dec.

Table 8 gives certain parameters for the various symmetrical weighted averages that have been mentioned previously. The column headed "Error" requires explanation. This is the error committed when the formula in question is used to "smooth" a polynomial of degree four. This naturally tends to increase with the number of terms in the formula. Both R_0 and R_3 tend to decrease with increasing number of terms. Though the "ideal" formulas have been derived to minimize R_3 , they tend to produce small values of R_0 as well. In only one instance (Vaughan) does a "name" formula have a smaller R_0 than the ideal formula of the same number of terms. The late Hubert Vaughan was an unusually keen analyst of MWA smoothing.

It may be mentioned in passing that some writers (e. g., Henderson 1938) call the reciprocal of R_0^2 the "weight" and the reciprocal of R_3 the (smoothing) "power."

9. PROOF OF EQUIVALENCE OF THE MATRIX AND INTERMEDIATE-VALUE APPROACHES

Though this proof involves only elementary mathematics, it is fairly long and complicated, and is therefore organized in the form of three lemmas and a theorem.

Let

$$p(z) = z^{m-s} - \sum_{j=1}^{m-s} p_j z^{m-s-j},$$

where $p(z)$ is the polynomial defined in Section 3 whose zeros are the zeros of $q(z)$ located inside the unit circle.

Lemma 9.1. The quantities h_j of (6.2) satisfy the recurrence

$$h_j = \sum_{l=1}^{m-s} p_l h_{j-l} \quad (9.1)$$

for all positive j .

Proof. In an annular region containing the unit circle, we have

$$h(z) q(z) = 1.$$

But, for a suitable (nonvanishing) constant c ,

$$q(z) = cp(z) p(z^{-1}). \quad (9.2)$$

In fact, $c = -q_{m-s}/p_{m-s}$. Therefore,

8. Parameters of the Symmetrical Weighted Averages Listed in Tables 2 and 3

Designation	Number of Terms	R_0	R_3	Error
Minimum- R_3 (Henderson's ideal):	5	.7045	.2735	-.0736 ⁴
	7	.5971	.1147	-.296 ⁴
	9	.5323	.0581	-.766 ⁴
	11	.4865	.0331	-1.576 ⁴
	13	.4515	.0204	-2.886 ⁴
	15	.4234	.0134	-4.856 ⁴
	17	.4002	.0095	-7.646 ⁴
	19	.3806	.0066	-11.46 ⁴
	21	.3636	.0048	-16.56 ⁴
	23	.3488	.0036	-23.16 ⁴
Macaulay	15	.4273	.01657	-4.526 ⁴
Spencer	15	.4389	.01659	-3.866 ⁴
Woolhouse	15	.4602	.0654	-5.46 ⁴
Hardy	17	.4059	.0105	$\frac{1}{12} \delta^2 - 3.706^4$
Higham	17	.4127	.0179	-6.46 ⁴
Karup	19	.4036	.0095	-7.86 ⁴
Andrews	21	.3707	.00628	-14.96 ⁴
Spencer	21	.3784	.00626	-12.66 ⁴
Hardy, wave-cutting	23	.3332	.0154	-48.86 ⁴
Vaughan A	23	.3415	.0050	-26.66 ⁴
Kenchington	27	.3202	.0031	-22.46 ⁴

$$cp(z) p(z^{-1}) h(z) = 1 .$$

(9.3)

Now, $[p(z)]^{-1}$ has an expansion in negative powers of z , with exponents not greater than $-m + s$, whose region of convergence contains the unit circle. Call it $b(z)$. Then,

$$cp(z^{-1}) h(z) = b(z) ,$$

from which it follows that (9.1) holds for all positive j , and the proof is complete.

Let D_{11} denote the (unknown) square submatrix of order $m - s$ in the upper left corner of D . Let $P = (p_{ij})$ be a matrix of $m - s$ rows and $2m - 2s$ columns defined by

$$p_{ij} = \begin{cases} 0 & \text{for } i > j \\ 1 & \text{for } i = j \\ -p_{j-i} & \text{for } 0 < j - i \leq m - s \\ 0 & \text{for } j - i > m - s . \end{cases}$$

Let P be partitioned in the form $[P_1 P_2]$, where P_1 and P_2 are square. Let $T = (t_{ij})$ denote the Toeplitz matrix of order $N - s$ defined by

$$t_{ij} = h_{i-j} .$$

In other words, T is D^{-1} under assumption (iii) of Section 6.

Lemma 9.2. $-T$ is nonsingular and equal to D^{-1} if D is completed by assigning

$$D_{11} = cP_1^T P_1 ,$$

together with a corresponding assignment of the square submatrix of order $m - s$ in the lower right corner of D .

Proof. Note that if we try to form the product DT , all elements of the product that do not involve the missing elements in the corners of D have the correct values (0 or 1). We shall focus on the upper left corner; similar considerations apply to the lower right corner. The lemma will be proved if it can be shown that the product DT is indeed the identity if D_{11} (and its counterpart at the lower right) is chosen in the manner indicated.

Let D_{12} denote the square submatrix of D of order $m - s$ immediately to the

right of D_{11} , let T_{11} be the submatrix corresponding to D_{11} in the upper left corner of T , and T_{12} the one immediately to its right. By symmetry the square submatrix of order $m - s$ immediately below T_{11} is T_{12}^T . The product DT will be the identity if D_{11} is such that

$$D_{11} T_{11} + D_{12} T_{12}^T = I \quad (9.4)$$

(and if a similar relation holds in the lower right corner).

It is easily verified that

$$D_{12} = c P_1^T P_2. \quad (9.5)$$

Let $\tilde{D}_{11} = (\tilde{d}_{ij})$ be a square matrix of order $m - s$ defined by

$$\tilde{d}_{ij} = q_{i-j}.$$

The reader will note that replacement of D_{11} by \tilde{D}_{11} (and a corresponding replacement in the lower right corner) would make D a Toeplitz matrix. It follows from the definition of $h(z)$ in (6.2) and the Toeplitz character of the matrices involved that

$$D_{12}^T T_{12} + \tilde{D}_{11} T_{11} + D_{12} T_{12}^T = I. \quad (9.6)$$

(The reader may think of the block immediately below D_{11} as moved up to the left of D_{11} , and the block T_{12} moved to a position immediately above T_{11} .) It is clear from (9.6) that (9.4) will be satisfied if

$$D_{12}^T T_{12} + \tilde{D}_{11} T_{11} = D_{11} T_{11}. \quad (9.7)$$

It is easily verified that

$$\tilde{D}_{11} = [P_2^T P_1^T] \begin{bmatrix} P_2 \\ P_1 \end{bmatrix} = P_2^T P_2 + P_1^T P_1.$$

Substitution of this result and (9.5) in the left member of (9.7) gives

$$c(P_2^T P_1 T_{12} + P_2^T P_2 T_{11} + P_1^T P_1 T_{11}). \quad (9.8)$$

But

$$P_1 T_{12} + P_2 T_{11} = P \begin{bmatrix} T_{12} \\ T_{11} \end{bmatrix} = 0$$

by Lemma 9.1. Thus (9.8) reduces to $cP_1^T P_1 T_{11}$, and (9.7) is satisfied if $D_{11} = cP_1^T P_1$. This completes the proof.

Let L denote the m by N submatrix of $I - G = K^T D K$ consisting of the first m rows, and let $A = (a_{ij})$ be the m by N matrix defined by

$$a_{ij} = \begin{cases} 0 & \text{for } i > j \\ 1 & \text{for } i = j \\ -a_{j-i} & \text{for } 0 < j - i \leq m \\ 0 & \text{for } j - i > m, \end{cases} \quad (9.9)$$

where the coefficients a_j were defined in (3.4). Let A_1 denote the square submatrix of A consisting of the first m columns.

Lemma 9.3.

$$L = cA_1^T A. \quad (9.10)$$

Proof. Let D_1 denote the submatrix of D consisting of the first m rows, and let K_{11} denote the square submatrix of order m in the upper left corner of K . Then it follows from the placement of zeros in K^T that

$$L = K_{11}^T D_1 K.$$

Let \hat{P} denote an m by $N - s$ matrix with the elements defined as in P (following the proof of Lemma 9.1). It is easily verified that

$$A = \hat{P}K.$$

Let \hat{P}_1 denote the square submatrix of \hat{P} consisting of the first m columns. Then

$$A_1^T = K_{11}^T \hat{P}_1^T.$$

Thus,

$$cA_1^T A = cK_{11}^T \hat{P}_1^T \hat{P}K.$$

But it follows from the proof of Lemma 9.2 that $c\hat{P}_1^T \hat{P} = D_1$, and so

$$cA_1^T A = K_{11}^T D_1 K = L,$$

as required for the proof of the lemma.

Theorem 9.1. The extension method of Section 3 and the matrix formulation of Section 6 are equivalent.

Proof. Let A_2 denote the submatrix of A consisting of the $(m+1)$ th to $(2m)$ th columns and let

$$\hat{A} = [A_1 \ A_2] .$$

Let $y^{(0)}$ denote the vector of the m intermediate values obtained from the observations by (3.5), let $y^{(1)}$ and $y^{(2)}$, respectively, denote the vectors of the first m observations and the $(m+1)$ th to $(2m)$ th observations, and let \hat{y} denote the vector consisting of $y^{(0)}$ followed by $y^{(1)}$. Then, the extension method requires $\hat{A}\hat{y} = 0$, or

$$A_1 y^{(0)} = -A_2 y^{(1)} , \quad (9.11)$$

Let $\hat{G} = (\hat{g}_{ij})$ be the square matrix of order m defined by $\hat{g}_{ij} = c_{i-j}$ (where the coefficients c_j were defined in (1.2)), and let G_{12} be the submatrix of G formed from the first m rows and the $(m+1)$ th to $(2m)$ th columns. Then the vector of the first m graduated values from the matrix formulation is, by Lemma 9.3,

$$y^{(1)} = cA_1^T (A_1 y^{(1)} + A_2 y^{(2)}) . \quad (9.12)$$

By the extension method, the corresponding vector is

$$G_{12}^T y^{(0)} + \hat{G} y^{(1)} + G_{12} y^{(2)} . \quad (9.13)$$

But, since $G = I - K^T DK$,

$$\hat{G} = I - c[A_2^T \ A_1^T] \begin{bmatrix} A_2 \\ A_1 \end{bmatrix} = I - c[A_1^T \ A_1 + A_2^T \ A_2] ,$$

and

$$G_{12} = -cA_1^T A_2 .$$

Thus, (9.13) reduces to

$$y^{(1)} = c[A_2^T A_1 y^{(0)} + A_1^T A_1 y^{(1)} + A_2^T A_2 y^{(1)} + A_1^T A_2 y^{(2)}] .$$

The substitution (9.11) reduces this to (9.12), as required.

We note in passing that the computational short cut involving extended values has an analogue in the case of Whittaker smoothing. Especially in actuarial literature, the Whittaker smoothing process is sometimes called the difference-equation method because the difference equation

$$u_x + (-1)^s \delta^{2s} u_x = y_x \quad (9.14)$$

holds for $x = A + s, A + s + 1, \dots, B - s$. It was pointed out by Aitken (1926) that (9.14) is satisfied for $x = A, A + 1, \dots, B$ if we introduce at each end of the data set s extrapolated values of both y_x and u_x satisfying the conditions

$$u_x = y_x \quad (x = A - j, x = B + j; j = 1, 2, \dots, s),$$

$$\Delta^s u_x = 0 \quad (x = A - j, x = B - j; j = 1, 2, \dots, s).$$

However, this observation is not helpful from a computational point of view. The attempt to utilize it merely increases the order of the linear system to be solved from N to $N + 2s$.

10. THE CHARACTERISTIC FUNCTION AND SCHOENBERG'S DEFINITION OF A SMOOTHING FORMULA

Schoenberg (1946) defined the characteristic function of the MWA (1.2) as

$$\phi(t) = \sum_{j=-m}^m c_j e^{ijt}. \quad (10.1)$$

For a symmetrical MWA this is a real function of the real variable t , and can be expressed in the alternative form

$$\phi(t) = \sum_{j=-m}^m c_j \cos jt.$$

It is periodic with period 2π and equal to unity for $t = 2\pi n$ for all integers n .

The effect of MWA's in eliminating or reducing certain waves has been noted (Elphinstone 1951, Hannan 1970). If the input to the smoothing process is a sine wave, which may be represented in the form

$$y_x = C \cos(rx + h), \quad (10.2)$$

it can be shown by simple algebraic manipulation that

$$u_x = y_x \phi(2\pi/P)$$

where $P = 2\pi/r$ is the period of y_x . Thus, if $\phi(2\pi/P) = 0$, the wave is annihilated by the smoothing process; the amplitude is severely reduced if it is close to zero. Thus MWA smoothing is related to the "filtering" processes considered by Wiener (1949) and others.

Schoenberg (1946) defined a smoothing formula as an MWA whose characteristic function $\phi(t)$ satisfies the condition

$$|\phi(t)| \leq 1 \quad (10.4)$$

for all t . Later (Schoenberg 1948, 1953) he suggested the stronger condition

$$|\phi(t)| < 1 \quad (0 < t < 2\pi) \quad (10.5)$$

C. Lanczos (see Schoenberg 1953) pointed out that condition (10.4) is obtained by requiring that every simple vibration (10.2) be diminished in amplitude by the transformation (1.2). The results of Section 6 of the present paper suggest an alternative definition of a smoothing formula. Using the subscript N to emphasize the fact that the order of G is the number of observations in the data set, we may say that (1.2) is a smoothing formula if

$$G_N^\infty = \lim_{n \rightarrow \infty} G_N^n \quad (10.6)$$

exists for all N . Schoenberg (1953, footnote 3) suggested a relationship between (10.4) and the conditions for existence of the infinite power of a matrix (Oldenburger 1940, Dresden 1942), but he did not elaborate the connection. We shall show that the existence of the limit (10.6) for all N is equivalent to a condition intermediate between (10.4) and (10.5). The following lemma will help to elucidate the situation.

Lemma 10.1. For a given τ in $(0, 2\pi)$, $\phi(\tau) = 1$ if and only if $q(\tau) = 0$.

Proof. From (3.1), (3.2), and (10.1) it follows that

$$\phi(t) = 1 - (-1)^s (2i \sin \frac{1}{2}t)^{2s} q(e^{it}) = 1 - (4 \sin^2 \frac{1}{2}t)^s q(e^{it}). \quad (10.7)$$

Since $\sin \frac{1}{2}\tau \neq 0$, the lemma is established.

There are two ways in which equality can hold in (10.4), namely $\phi(t) = 1$ and $\phi(t) = -1$, and the situation is different in the two cases. Lemma 10.1 shows that if $\phi(t) = 1$, $q(z)$ has a zero on the unit circle and consequently G_N is not defined. No such problem arises if $\phi(t) = -1$. We are therefore led to the intermediate condition

$$-1 \leq \phi(t) < 1 \quad (0 < t < 2\pi), \quad (10.8)$$

which we shall show to be equivalent to the existence of (10.6).

Lemma 10.2. If (10.8) holds, D is positive definite.

Proof. From (9.2) we obtain

$$q(1) = c[p(1)]^2.$$

It follows from (10.7) and (10.8) that $q(e^{it})$ is positive for $0 < t < 2\pi$. Since it is a continuous function of t , it is nonnegative for $t = 0$: that is, $q(1)$ is nonnegative. By the definition of $p(z)$, $p(1) \neq 0$, and $c = -q_{m-s}/p_{m-s}$ does not vanish. Therefore $q(1)$ is positive and c is positive.

Let the expansion of $b(z)$ of Section 9 be given by

$$b(z) = \sum_{j=m-s}^{\infty} b_j z^{-j}. \quad (10.9)$$

It follows from (9.3) that on the unit circle

$$b(z) b(z^{-1}) = ch(z). \quad (10.10)$$

Substitution of (10.9) gives

$$\sum_{l=m-s}^{\infty} b_l b_{l+j} = ch_j \quad (j = \dots, -1, 0, 1, \dots). \quad (10.11)$$

We note that the coefficients b_j satisfy the difference equation

$$b_j = \sum_{l=1}^{m-s} p_l b_{j-l} \quad (j = m-s+1, m-s+2, \dots).$$

If r_1, r_2, \dots, r_{m-s} are the zeros of $p(z)$, it follows that

$$b_j = \sum_{k=1}^{m-s} a_k r_k^j$$

for some constant coefficients a_k . Therefore the series in the left member of (10.11) is absolutely convergent.

Let $B = (b_{ij})$ denote the matrix of $N - s$ rows and a denumerable infinity of columns given by

$$b_{ij} = \begin{cases} 0 & (i > j) \\ b_{m-s+j-i} & (i \leq j) \end{cases} . \quad (10.12)$$

It follows from (10.11) and (6.1) that

$$D^{-1} = cBB^T . \quad (10.13)$$

The structure of the right member of (10.13) shows that D^{-1} is nonnegative definite; since it is nonsingular, it is positive definite, and consequently D is positive definite, as required.

Let $\psi(t)$ be defined by

$$\psi(t) = 1 - \phi(t) .$$

Then (10.8) is equivalent to

$$0 < \psi(t) \leq 2 \quad (0 < t < 2\pi) . \quad (10.14)$$

Let ψ_{\max} denote the maximum value of $\psi(t)$, and let $\tilde{A} = (a_{ij})$ be the square matrix of order N defined by (9.9). Let $|v|$ denote the Euclidean norm of a vector v .

Lemma 10.3. If v is any vector of N real components and $\psi(t)$ is positive in $(0, 2\pi)$,

$$|\tilde{A}v|^2 / |v|^2 \leq c^{-1} \psi_{\max} . \quad (10.15)$$

Proof. Let v be an arbitrary vector of N real components, with j th component v_j , and let

$$v(t) = \sum_{j=1}^N v_j e^{ijt} \quad (10.16)$$

be the characteristic function of v . Then, if $a(z)$ is the polynomial defined by (3.4),

$$a(e^{it}) v(t) = \sum_{j=1}^{N+m} w_j e^{ijt} ,$$

where, for $j \geq m$, w_j is the $(j - m)$ th component of $\tilde{A}v$. Moreover, it follows from (3.4), (9.2), and (10.7) that

$$\psi(t) = c(a(e^{it}) a(e^{-it})) . \quad (10.17)$$

By Parseval's formula (Schoenberg 1946)

$$|v|^2 = \frac{1}{2\pi} \int_0^{2\pi} |v(t)|^2 dt . \quad (10.18)$$

Similarly, in view of (10.17),

$$|\tilde{A}v|^2 \leq \frac{1}{2\pi} \int_0^{2\pi} |a(e^{it})|^2 |v(t)|^2 dt = \frac{c^{-1}}{2\pi} \int_0^{2\pi} \psi(t) |v(t)|^2 dt . \quad (10.19)$$

From (10.18) and (10.19), (10.15) follows.

It is easily verified that the symmetric matrix $\tilde{A}^T \tilde{A}$ is identical with $F = I - G$ except for the elements of the square submatrix in the lower right corner. In fact, the elements in the lower right corner of $\tilde{A}^T \tilde{A}$ are such that the entire matrix becomes a Toeplitz matrix if the first m rows and the first m columns are deleted. It follows that F can be obtained from $c\tilde{A}^T \tilde{A}$ by subtracting a square matrix of order N whose elements are all zero except for the square submatrix of order m in the lower right corner.

In fact, if A_1 and A_2 are defined as in the proof of Theorem 9.1, it is easily verified that the square submatrix of order m in the lower right corner of $c\tilde{A}^T \tilde{A}$ is given by

$$c[A_2^T A_1^T] \begin{bmatrix} A_2 \\ A_1 \end{bmatrix} = c(A_2^T A_2 + A_1^T A_1) .$$

while the corresponding submatrix of F is $cA_1^T A_1$. If, therefore, \bar{A} is defined as the square matrix of order N having A_1 in the lower right corner and zeros everywhere else, we have

Lemma 10.4.

$$F = I - G = c(\tilde{A}^T \tilde{A} - \bar{A}^T \bar{A}) .$$

Before stating the theorem that is the main result of this section, we point out (Oldenberger 1940, Dresden 1942) that, for a given matrix C , $\lim_{n \rightarrow \infty} C^n$ exists if and only

if either all eigenvalues of C lie within the unit circle, or else 1 is a simple zero of the minimum polynomial of C and all other eigenvalues lie within the unit circle. As multiplication by G leaves unchanged vectors whose components are successive equally spaced ordinates of a polynomial of degree $s - 1$ or less, it is clear that 1 is an eigenvalue. As G is symmetric, all its eigenvalues are real, and all zeros of its minimum polynomial (including 1) are simple. Therefore the limit (10.6) exists if and only if all eigenvalues of C other than 1 are strictly between -1 and 1.

Theorem 10.1. The limit $\underset{N}{G}^\infty$ exists for all N if and only if the characteristic function $\phi(t)$ satisfies the condition

$$-1 \leq \phi(t) < 1 \quad (0 < t < 2\pi). \quad (10.8)$$

Proof. Let (10.8) hold, and recall that (10.8) is equivalent to (10.14). Now we shall consider a particular value of N and, for convenience, drop the subscript N . The eigenvalues of F are obtained by subtracting from 1 those of G . We need to show, therefore, that the nonzero eigenvalues of F are positive and do not exceed 2.

Since $F = K^T D K$ and D is positive definite by Lemma 10.2, F is nonnegative definite. Therefore its nonzero eigenvalues are positive. Let v be an arbitrary nonzero real vector and consider the Rayleigh quotient,

$$r = \frac{\frac{v^T F v}{v^T v}}{c} = c \frac{\frac{v^T \tilde{A}^T \tilde{A} v}{v^T v}}{c} = c \frac{\frac{v^T \tilde{A}^T \tilde{A} v}{v^T v}}{c}$$

by Lemma 10.4. Since $\tilde{A}^T \tilde{A}$ is nonnegative definite, we have

$$r \leq c |\tilde{A} v|^2 / |v|^2 \leq \psi_{\max}$$

by Lemma 10.3. Therefore (10.14) gives $r \leq 2$. Since the spectral radius is the maximum value of the Rayleigh quotient, this completes the first part of the proof.

We shall prove the converse by showing that if (10.8) fails, the limit (10.6) does not exist for some N . There are two ways in which (10.8) can fail. Either $\phi(t)$ may be equal to 1, or $|\phi(t)|$ may exceed 1. In the former case, as was pointed out immediately following Lemma 10.1, G is not defined.

We consider therefore the case in which $|\phi(t)| > 1$ for some $t = \tau$, and let v be the N -vector whose j th component is

$$v_j = \exp[i\tau(j - \frac{N+1}{2})]. \quad (10.20)$$

Using an asterisk to denote the conjugate transpose, we have $\bar{v}_j^* v_j = 1$, and therefore

$$\bar{v}^* v = N. \quad (10.21)$$

Except for the first m and the last m components we have

$$(Gv)_j = \phi(\tau) v_j,$$

and so

$$\bar{v}^* Gv = \phi(\tau)(N - 2m) + C, \quad (10.22)$$

where C denotes the contribution of the first m and the last m components. Because of the symmetry of both the matrix elements and the vector components, C is real. Since all the vector components have absolute value 1, an upper bound to C is the sum of the absolute values of the elements in the first m and the last m rows of G . Call this C_1 . We recall that C_1 does not depend on N .

Now choose N sufficiently large so that

$$N > \frac{C_1 + 2m|\phi(\tau)|}{|\phi(\tau)| - 1}.$$

Then

$$N[|\phi(\tau)| - 1] > C_1 + 2m|\phi(\tau)|,$$

and it follows that

$$(N - 2m)|\phi(\tau)| > N + |C|.$$

Consequently,

$$|(N - 2m)\phi(\tau) + C| > N,$$

and therefore, by (10.21) and (10.22),

$$|r| = \left| \frac{\bar{v}^* Gv}{\bar{v}^* v} \right| > 1.$$

It follows that the spectral radius of G is greater than unity, and the proof is complete.

It is easily verified that G^∞ , when it exists, is the orthogonal projector on the eigenspace of G associated with the eigenvalue 1, that is, the space of N -vectors whose components are successive equally spaced ordinates of polynomials of degree $s - 1$ or less.

11. SMOOTHING FORMULAS IN THE STRICT SENSE AND AN OPTIMAL PROPERTY

At an early stage of the investigations underlying this paper I was trying to explain the natural extension of the MWA graduation to my colleague, I. J. Schoenberg, whose work plays such an important role therein, and he asked me (I thought with a slight show of impatience) "What does it minimize?" My answer was that it doesn't minimize anything, but is just a natural way of extending the MWA graduation to the ends of the data. This was too simplistic an answer, for we shall now show that it does in fact minimize "something."

In a slightly more general form of the Whittaker smoothing method (Greville 1957) one minimizes the sum of the squares of the departures of the smoothed values from the observed values plus a designated positive definite quadratic form in the s th differences of the smoothed values. In other words, one minimizes

$$(u - y)^T (u - y) + (Ku)^T HKu ,$$

where H is a given positive definite matrix of order $N - m$. Minimization of this expression leads to the equation

$$(I + K^T HK)u = y ,$$

which has a unique solution for u since $I + K^T HK$ is positive definite. I showed (Greville 1957) that this graduation method has the interesting property that if roughness (opposite of smoothness) is measured by the term $(Ku)^T HKu$, smoothness is always increased by the graduation. By Theorem 5.22 of Noble (1969),

$$(I + K^T HK)^{-1} = I - K^T (H^{-1} + K K^T)^{-1} K .$$

The last expression is of the form (5.6) and suggests that the use of an MWA with the natural extension might be regarded as a generalized Whittaker smoothing process if

$$D = (H^{-1} + K K^T)^{-1} .$$

Solving for H gives

$$H = (D^{-1} - KK^T)^{-1} . \quad (11.1)$$

We are led to inquire, therefore, under what conditions an MWA is such that the right member of (11.1) is positive definite. Clearly H is positive definite if and only if the Toeplitz matrix

$$H^{-1} \approx D^{-1} - KK^T \quad (11.2)$$

is positive definite.

Schoenberg (1946, p. 53) remarks that it is desirable for an efficient smoothing formula, one that achieves adequate smoothness without producing unnecessarily large departures from the observed values, to have its characteristic function satisfy the stronger condition

$$0 \leq \phi(t) \leq 1 .$$

This remark seems to have been little noted in the years since its publication. We shall call an MWA a smoothing formula in the strict sense if its characteristic function satisfies the condition

$$0 \leq \phi(t) < 1 \quad (0 < t < 2\pi) , \quad (11.3)$$

and we shall show that (11.2) is positive definite for all N if and only if (11.3) holds.

Theorem 11.1. Let (10.8) hold. Then $Q = D^{-1} - KK^T$ is positive definite for all N if and only if the MWA is a smoothing formula in the strict sense.

Proof. Let (11.3) hold, let v be an arbitrary nonzero real N -vector, and consider the Rayleigh quotient,

$$r = \frac{v^T Q v}{v^T v} = (c|B^T v|^2 - |K^T v|^2)/|v|^2 , \quad (11.4)$$

where B is given by (10.12). Let $V(t)$ be defined by (10.16). Then, by Parseval's formula,

$$\begin{aligned} |B^T v|^2 &= \frac{1}{2\pi} \int_0^{2\pi} |e^{-(m-s)it} b(e^{-it})|^2 |V(t)|^2 dt \\ &= \frac{c^{-1}}{2\pi} \int_0^{2\pi} |h(e^{it})||V(t)|^2 dt \end{aligned} \quad (11.5)$$

by (10.10). Moreover, again by Parseval's formula,

$$|\mathbf{k}^T \mathbf{v}|^2 = \frac{1}{2\pi} \int_0^{2\pi} (4\sin^2 \frac{1}{2}t)^s |v(t)|^2 dt . \quad (11.6)$$

It was shown in the proof of Lemma 10.2 that $q(e^{it})$ is positive, and therefore $h(e^{it}) = [q(e^{it})]^{-1}$ is positive, for $0 < t < 2\pi$. By means of (11.5), (11.6), and (10.7), (10.4) gives

$$\begin{aligned} r &= \frac{1}{2\pi} \int_0^{2\pi} [h(e^{it}) - (4\sin^2 \frac{1}{2}t)^s] |v(t)|^2 dt \\ &= \frac{1}{2\pi} \int_0^{2\pi} h(e^{it}) \phi(t) |v(t)|^2 dt > 0 , \end{aligned}$$

since any zeros of $\phi(t)$ constitute a set of measure zero. Since r is positive for arbitrary nonzero v , Q is positive definite.

To prove the converse we shall suppose that (11.3) does not hold and show that r is negative for some N and some v . Because of the hypothesis that (10.8) holds, $\phi(t)$ is less than 1. We suppose, therefore, that $\phi(t) < 0$ for some $t = \tau$ in $(0, 2\pi)$. Let v be given by (10.20). By an argument similar to that used in the proof of Lemma 10.2, it is easily shown that the series (6.2) for $h(z)$ is absolutely convergent for $z = 1$. Thus, for any small positive quantity ϵ , there exists a positive integer M such that

$$\sum_{j=M+1}^{\infty} |h_j| < \frac{1}{2} \epsilon .$$

Thus, for $N > 2M$, the j th component of Qv , for $j = M + 1, M + 2, \dots, N - M$, is v multiplied by a quantity less than

$$h(e^{i\tau}) + \epsilon - (4\sin^2 \frac{1}{2}\tau)^s . \quad (11.7)$$

By (10.7).

$$h(e^{i\tau}) - (4\sin^2 \frac{1}{2}\tau)^s = h(e^{i\tau}) \phi(\tau) . \quad (11.8)$$

It was shown in the proof of Lemma 10.2 that (10.8) implies $h(e^{i\tau}) > 0$. Thus, (11.8) is negative. Choose ϵ sufficiently small so that (11.7) is negative.

As in the proof of Theorem 10.1, we find that $v^* v = N$, while

$$v^* Qv \leq (N - 2M) [h(e^{i\tau}) \phi(\tau) + \epsilon] + C, \quad (11.9)$$

where C denotes the contribution of the first M and the last M rows. As in the proof of Theorem 10.1, it follows from symmetry that C is real. Let

$$n = \sum_{j=-\infty}^{\infty} |h_j - (-1)^{s-j} \binom{2s}{s-j}|,$$

where $\binom{2s}{j}$ is understood to vanish for negative j or $j > 2s$. Then

$$C \leq |C| \leq 2Mn,$$

since all components of v have absolute value unity. Now take

$$N > \frac{2M[h(e^{i\tau}) \phi(\tau) + \epsilon] - n}{h(e^{i\tau}) \phi(\tau) + \epsilon}. \quad (11.10)$$

Note that both numerator and denominator of the right member of (11.10) are negative.

From (11.10) we obtain

$$(N - 2M)[h(e^{i\tau}) \phi(\tau) + \epsilon] < -2Mn,$$

and (11.9) then gives $v^* Qv < 0$, so that Q is not positive definite. This completes the proof of the theorem.

It is easy to construct an MWA that is a smoothing formula in the strict sense. A trivial example is the formula

$$u_x = \frac{1}{17} (-y_{x-2} + 4y_{x-1} + 11y_x + 4y_{x+1} - y_{x+2}).$$

However, none of the weighted averages in general use fall in this class. In particular, using the properties of Jacobi polynomials, I have shown elsewhere (Greville 1966) that the characteristic functions of all minimum- R_s averages assume negative values in $(0, 2\pi)$. Thus no such formula is a smoothing formula in the strict sense.

There is, however, one family of moving averages, mentioned in the literature but not in general use, that are smoothing formulas in the strict sense. Elsewhere (Greville 1966) I have considered the limiting case of the minimum- R_s formulas as s tends to infinity. In finite-difference form, the minimum- R_∞ MWA of $2m + 1$ terms, exact for the degree $2s - 1$, is

$$u_x = \mu^{2(m-s+1)} \sum_{j=0}^{s-1} (-4)^{-j} \binom{m-s+j}{j} \delta^{2j} y_x ,$$

where the operator μ is defined by

$$\mu f(x) = \frac{1}{2}[f(x + \frac{1}{2}) + f(x - \frac{1}{2})] ,$$

so that $\mu^2 = 1 + \frac{1}{4}\delta^2$. The characteristic function is

$$\phi(t) = (\cos \frac{1}{2}t)^{m-s+1} \sum_{j=0}^{s-1} \binom{m-s+j}{j} \sin^{2j} \frac{1}{2}t ,$$

which is nonnegative, with a single zero of order $m - s + 1$ at $t = \pi$.

By a tour de force it is possible to show that, for an MWA that is not a smoothing formula in the strict sense, but whose characteristic function satisfies (10.8), the natural extension does nevertheless "minimize something." For the given MWA, let $-\rho$ denote the minimum value of $\phi(t)$, and let γ be chosen so that $0 < \gamma \leq (1 + \rho)^{-1}$. Then $1 - \gamma(1 + \rho) \geq 0$. Let a modified MWA, \tilde{u}_x be obtained by taking

$$\tilde{u}_x = \gamma u_x + (1 - \gamma) y_x . \quad (11.11)$$

Clearly this is an MWA of the form (1.2),

$$\tilde{u}_x = \sum_{j=-m}^m \tilde{c}_j y_{x-j} ,$$

with $\tilde{c}_0 = \gamma c_0 + 1 - \gamma$ and $\tilde{c}_j = \gamma c_j$ for $j \neq 0$. The modified MWA is a smoothing formula in the strict sense, and its graduation matrix is $\tilde{G} = I - K^T \tilde{D} K$, with $\tilde{D} = \gamma D$.

The modified graduation minimizes the quantity

$$(\tilde{u} - y)^T (\tilde{u} - y) + (K\tilde{u})^T \tilde{H} K\tilde{u} , \quad (11.12)$$

where

$$\tilde{H} = (\tilde{D}^{-1} - K K^T)^{-1} \quad (11.13)$$

is positive definite. Using (11.11) and (11.13) to express (11.12) in terms of the original graduated values, we find that the quantity minimized is

$$(u - y)^T (u - y) + [u + (\gamma^{-1} - 1)y]^T K^T \hat{Q}^{-1} K [u + (\gamma^{-1} - 1)y] . \quad (11.14)$$

where

$$\hat{Q} = \gamma^{-1} D^{-1} - K K^T$$

is positive definite. Thus, the total smoothing operation including the "tails," based on an MWA that is a smoothing formula, but not in the strict sense, does in fact minimize the expression (11.14). Using statistical terminology, this expression may therefore be regarded as a "loss function," but in that context is difficult to interpret and justify in practical terms.

ACKNOWLEDGMENTS

This paper has benefited from discussions with many persons, but I wish to thank especially D. R. Schuette for invaluable help with the computations, and J. M. Hoem and W. F. Trench for their careful reading of the manuscript, which led to significant improvements. I am solely responsible for any errors that may be found.

REFERENCES

- Aitken, A. C. (1926), "The Accurate Solution of the Difference Equation Involved in Whittaker's Method of Graduation, and its Practical Application," Transactions of the Faculty of Actuaries, 11, 31-9.
- Andrews, George H., and Nesbitt, Cecil J. (1965), "Periodograms of Graduation Operators," Transactions of the Society of Actuaries, 17, 1-27.
- Benjamin, B., and Haycocks, H. W. (1970), The Analysis of Mortality and Other Actuarial Statistics, Cambridge: Cambridge University Press.
- De Forest, Erastus L. (1873), "On Some Methods of Interpolation Applicable to the Graduation of Irregular Series, Such as Tables of Mortality, &c., &c.," Smithsonian Report 1871, 275-339.
- (1875), "Additions to a Memoir on Methods of Interpolation Applicable to the Graduation of Irregular Series," Smithsonian Report 1873, 319-49.
- (1876), Additions to a Memoir on Methods of Interpolation Applicable to the Graduation or Adjustment of Irregular Series of Observed Numbers, New Haven: Tuttle, Morehouse, and Taylor Co.
- (1877), "On Adjustment Formulas," The Analyst, 4, 79-86 and 107-13.
- Dresden, Arnold (1942), "On the Iteration of Linear Homogeneous Transformations," Bulletin of the American Mathematical Society, 48, 577-9.
- Elphinstone, M. D. W. (1951), "Summation and Some Other Methods of Graduation -- the Foundations of Theory," Transactions of the Faculty of Actuaries, 20, 15-77.
- Greville, T. N. E. (1947), "Actuarial Note: Adjusted Average Graduation Formulas of Maximum Smoothness," Record of the American Institute of Actuaries, 36, 249-64.
- (1948), "Actuarial Note: Tables of Coefficients in Adjusted Average Graduation Formulas of Maximum Smoothness," Record of the American Institute of Actuaries, 37, 11-30.
- (1957), "On Smoothing a Finite Table: a Matrix Approach," Journal of the Society for Industrial and Applied Mathematics, 5, 137-54.
- (1966), "On Stability of Linear Smoothing Formulas," SIAM Journal on Numerical Analysis, 3, 157-70.

(1974a), Part 5 Study Notes, Graduation (1974 edition), Chicago: Education and Examination Committee of the Society of Actuaries.

(1974b), "On a Problem of E. L. De Forest in Iterated Smoothing," SIAM Journal on Mathematical Analysis, 5, 376-98.

Hannan, E. J. (1970), Multiple Time Series, New York: John Wiley and Sons.

Hardy, George F. (1909), The Theory of the Construction of Tables of Mortality and of Similar Statistical Tables in Use by the Actuary, London: Institute of Actuaries.

Henderson, Robert (1916), "Graduation by Adjusted Average," Transactions of the Actuarial Society of America, 17, 43-48.

(1924), "A New Method of Graduation," Transactions of the Actuarial Society of America, 25, 29-40.

(1938), Mathematical Theory of Graduation, New York: Actuarial Society of America.

Macaulay, Frederick R. (1931), The Smoothing of Time Series, New York: National Bureau of Economic Research.

Maclean, Joseph B. (1913), "Graduation by the Summation Method. Some Elementary Notes," Transactions of the Actuarial Society of America, 14, 256-76.

Miller, Morton D. (1946), Elements of Graduation, New York and Chicago: Actuarial Society of America and American Institute of Actuaries.

Noble, Ben (1969), Applied Linear Algebra, Englewood Cliffs, N. J.: Prentice-Hall, Inc.

Oldenburger, Rufus (1940), "Infinite Powers of Matrices and Characteristic Roots," Duke Mathematical Journal, 6, 357-61.

Schiaparelli, G. V. (1866), "Sul Modo di Ricavare la Vera Espressione delle Leggi della Natura dalle Curve Empiriche," appendix to Effemeridi Astronomiche di Milano per l'Anno 1867, Milan.

Schoenberg, I. J. (1946). "Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions," Quarterly of Applied Mathematics, 4, 45-99 and 112-41.

(1948), "Some Analytical Aspects of the Problem of Smoothing," in Studies and Essays Presented to R. Courant on his 60th Birthday, New York: Interscience Publishers.

- (1953), "On Smoothing Operations and Their Generating Functions," Bulletin of the American Mathematical Society, 59, 199-230.
- Sheppard, W. F. (1913), "Reduction of Errors by Means of Negligible Differences," Proceedings of the Fifth International Congress of Mathematicians, 2, 348-84.
- Shiskin, Julius, Young, Allan H., and Musgrave, John C. (1967), The X-11 Variant of the Census Method II Seasonal Adjustment Program, Washington: U. S. Department of Commerce, Bureau of the Census.
- Trench, William F. (1974), "Inversion of Toeplitz Band Matrices," Mathematics of Computation, 28, 1089-95.
- Vaughan, Hubert (1933), "Summation Formulas of Graduation with a Special Type of Operator," Journal of the Institute of Actuaries, 64, 428-48.
- Whittaker, E. T. (1923), "On a New Method of Graduation," Proceedings of the Edinburgh Mathematical Society, 41, 63-75.
- Wiener, Norbert (1949), Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications, New York: John Wiley and Sons.
- Wolfenden, Hugh H. (1925), "On the Development of Formulae for Graduation by Linear Compounding, with Special Reference to the Work of Erastus L. De Forest," Transactions of the Actuarial Society of America, 26, 81-121.
- (1942), The Fundamental Principles of Mathematical Statistics, Toronto: Macmillan Co. of Canada, Ltd.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 1786	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MOVING-WEIGHTED-AVERAGE SMOOTHING EXTENDED TO THE EXTREMITIES OF THE DATA		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T.N.E. Greville		8. CONTRACT OR GRANT NUMBER(s) DAAG29-75-C-0024
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of 610 Walnut Street Wisconsin Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 2 - Other Mathematical Methods
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P. O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE August 1977
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 52
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		15. SECURITY CLASS. (of this report) UNCLASSIFIED
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Smoothing, Toeplitz matrix, Laurent series, Moving weighted average		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A natural method for extending moving-weighted-average smoothing to the extremities of the data is developed.		